양자 컴퓨터 기반 문자열 간 사이먼 합동 판별 알고리즘

김성민, 한요섭

연세대학교 컴퓨터과학과

rena_rio@yonsei.ac.kr
emmous@yonsei.ac.kr

문제 정의

문자열 w에 대해 **부분열 함수** $w: \Sigma^* \rightarrow \{0,1\}$를 정의한다.

$$w(x) = 1 \Leftrightarrow x\text{는 } w\text{의 부분열}$$

정수 k에 대해 두 문자열 w_1, w_2이 **사이먼 합동**을 만족하면

$$\forall x \in \Sigma^{\leq k}: w_1(x) = w_2(x)$$

Q. 정수 k에 대해 두 문자열 w_1, w_2이
사이먼 합동($w_1 \sim_k w_2$)을 만족하는지 판별하시오.

예시)

$$aabb \sim_2 baba \quad \because \left\{ \begin{array}{l} aabb(aa) = baba(aa) = 1 \\ aabb(ab) = baba(ab) = 1 \\ \quad\quad\quad \vdots \\ aabb(\lambda) = baba(\lambda) = 1 \end{array} \right.$$

기존 연구

- Barker et al., "Scattered Factor-Universality of Words."

  전통적 컴퓨터로 사이먼 합동 판별 선형 시간 알고리즘 설계

- Kim et al., "On Simon's Congruence Closure of a String."

  문자열-정규언어 간 사이먼 합동 판별은 **NP-완전**함을 보임

- Kim et al., "On the Simon's Congruence Neighborhood of Langaugues."

  두 정규언어 간 사이먼 합동 판별은 **PSPACE-완전**함을 보임

기존 접근의 한계

- 길이 k이하 문자열의 개수는 **지수적으로 증가**
  - 모든 문자열에 대해 부분열 함숫값을 계산하고 비교하면 **다항 시간 내 해결 불가**

- 부분열의 성질을 이용한, **단일 문자열에 특화**된 판별 알고리즘 위주로 연구 진행
  - 문자열-언어, 언어-언어 간 사이먼 합동 판별 문제에 **확장 사용 불가**

연구의 필요성

- **양자적 접근**을 통한 빠른 부분열 함숫값 동일성 판단
  - 병렬 계산이 어려운 전통적 컴퓨터의 **한계 극복**
  - **양자 중첩 상태** 및 **양자 얽힘 현상** 활용
  - 양자 컴퓨터를 이용해 사이먼 합동을 판별하는 **첫 시도**

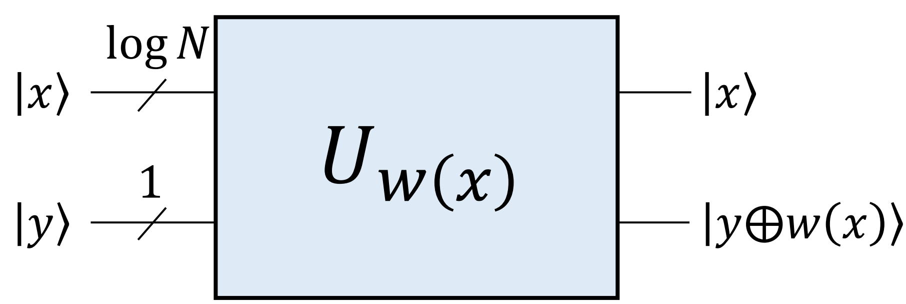
- 단일 문자열에 국한되지 않은 성질 활용
  - 사이먼 합동 **정의를 그대로 사용**하는 판별 알고리즘
  - 문자열-언어, 언어-언어 간 사이먼 합동 판별 문제에 **확장이 용이**

부분열 판별 양자 알고리즘 - 1

- $\frac{(|\Sigma|^{k+1}-1)}{|\Sigma|-1}$보다 작지 않은 최소 2의 승수를 N이라 둠.
  - $\log N$개 큐빗을 이용해 모든 $x \in \Sigma^{\leq k}$를 표현 가능
- x를 입력으로 받아 $w(x)$를 출력하는 알고리즘 $U_{w(x)}$ 설계
- 양자 알고리즘은 **유니타리 성질**을 만족해야 함

결과 중첩 상태로부터 입력 중첩 상태를 역산 가능

  - 입력 x값을 보존하는 알고리즘 설계
  - 추가 큐빗 y를 입력으로 받아 함숫값 출력



- 여러 $|x\rangle$가 양자 중첩 상태로 입력되더라도 출력은 각 $|x\rangle$에 대해 **독립적으로 계산**되어 양자 중첩 상태가 출력됨

$$\sum_{x=0}^{N-1} \frac{1}{\sqrt{N}} |x\rangle|y\rangle \xrightarrow{U_{w(x)}} \sum_{x=0}^{N-1} \frac{1}{\sqrt{N}} |x\rangle|y \oplus w(x)\rangle$$

N개의 상태가 균일 확률로 중첩된 입력

계수와 확률을 유지한 채로 독립적으로 계산됨

부분열 판별 양자 알고리즘 - 2

알고리즘 1: 부분열 판별 $U_{w(x)}$

1. 양자 계수기(counter) 레지스터 c_1, c_2를 준비하여 0으로 초기화한다.
2. $x[c_1] = w[c_2]$일 경우, c_1에 저장된 값을 1만큼 올린다.
3. c_2에 저장된 값을 1만큼 올린다.
4. c_2에 저장된 값이 w의 길이보다 작을 경우, 2.로 돌아간다.
5. c_1에 저장된 값이 k일 경우, y를 반전한다.

- 알고리즘에 사용된 모든 연산은 유니타리 연산으로 구현 가능
  - 덧셈 연산 (2., 3.)
  - 조건 연산 (4., 5.)
  - 비교 연산 (2., 4., 5.)
  - 큐빗 반전 연산 (5.)
- x를 문자열 단위가 아닌 문자 단위로 인코딩할 경우…
  - 입력으로 $\log N$개 큐빗이 아닌 $k\lceil\log|\Sigma|\rceil$개 큐빗 필요
  - 2.의 문자 접근 구현이 비교 연산으로 구현 가능

보조정리 1

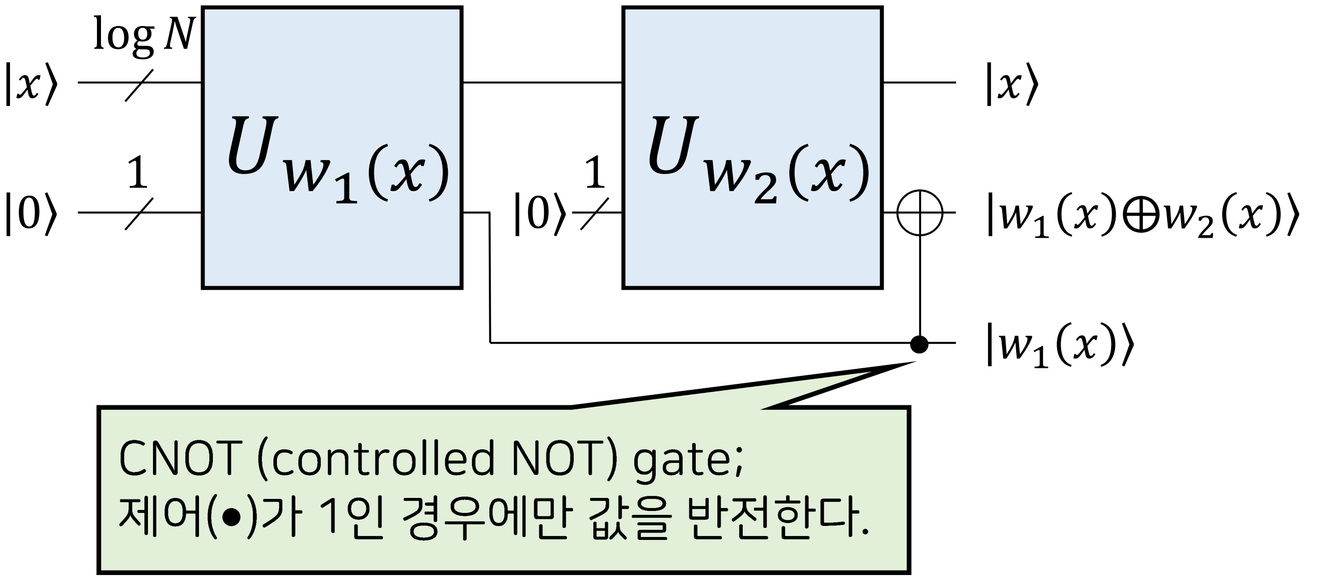
부분열 판별 알고리즘 $U_{w(x)}$ 은 $\sum_{x=0}^{N-1}\frac{1}{\sqrt{N}}|x\rangle|y\rangle$ 중첩 상태를 $\sum_{x=0}^{N-1}\frac{1}{\sqrt{N}}|x\rangle|y\oplus w(x)\rangle$로 변환하는 유니타리 연산이다.

균일 중첩 상태인 $\sum_{x=0}^{N-1}\frac{1}{\sqrt{N}}|x\rangle$는 0으로 초기화된 큐빗 $\log N$개를 하다마드(Hadamard) 게이트에 통과시켜 얻을 수 있다.

사이먼 합동 판별 양자 알고리즘 - 1

사이먼 합동 판별 회로 - 1

앞서 정의한 알고리즘 1을 두 번 활용한다.



- x가 w_1과 w_2 모두의 부분열일 경우 $|w_1(x) \oplus w_2(x)\rangle$는 $|0\rangle$이다.

정의

두 문자열 w_1과 w_2에 대해 **사이먼 거리**(Simon distance)는 길이 k 이하 문자열 중 $w_1(x) \neq w_2(x)$를 만족하는 x의 개수이다. 사이먼 거리는 $d_k(w_1, w_2)$로 나타낸다.

- 사이먼 합동 판별 회로 1에 균일 중첩 상태 $\sum_{x=0}^{N-1} \frac{1}{\sqrt{N}} |x\rangle$를 입력했을 경우 $|w_1(x) \oplus w_2(x)\rangle$이 $|1\rangle$**으로 관측될 확률**은 다음과 같다:

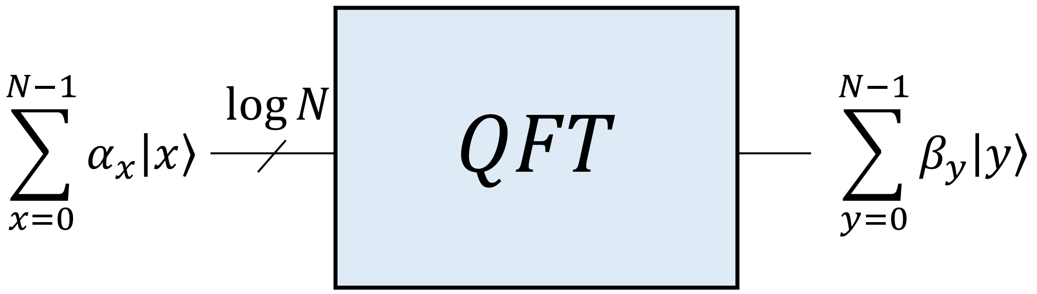
$$\frac{d_k(w_1, w_2)}{N}$$

사이먼 합동 판별 양자 알고리즘 - 2

$d_k(w_1, w_2)$가 0은 아니지만 0으로 수렴할 경우 $|w_1(x) \oplus w_2(x)\rangle$ 큐빗 관측만으로 사이먼 합동을 빠르게 판별하기는 어려움.

양자 푸리에 변환(Quantum Fourier Transform, QFT)

모든 $x = 0, 1, \ldots, N-1$에 대해 α_x는 주기성을 가지며 제곱해서 모두 합하면 1인 복소수라고 하자.



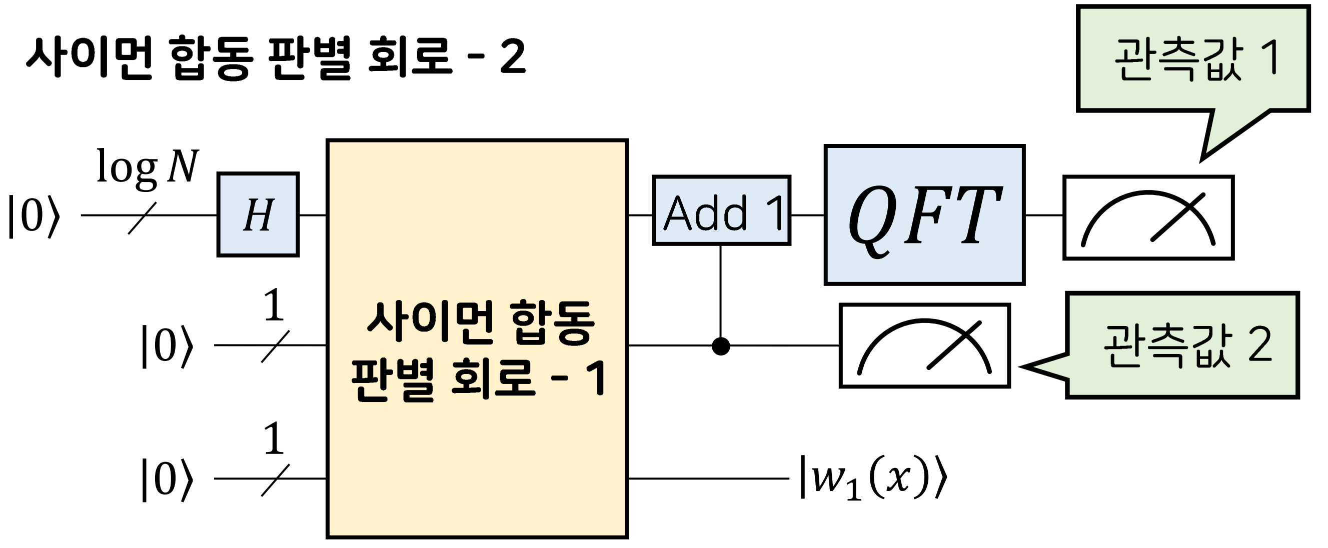
- 모든 $y = 0, 1, \ldots, N-1$에 대해 y가 α_x의 **진동수의 정수배**에 가까울수록 β_y값이 높다.
- 입력 α_x가 x에 대해 모두 균등하다면 $\beta_0 = 1$이고 $y \neq 0$인 모든 β_y는 0이다.

QFT를 활용한 $w_1 \nsim_k w_2$ 검출 확률 증가

- 사이먼 합동 판별 회로 – 1을 통해 얻은 $|w_1(x) \oplus w_2(x)\rangle$값이…
  - 모두 0일 경우, $|x\rangle$를 균일 중첩 상태로 유지한다.
  - 1이 있는 경우, $|x\rangle$의 **균일 중첩 상태를 흐트러뜨려** QFT 출력에서 0 이외의 값이 관측될 수 있게 한다.
- $|w_1(x) \oplus w_2(x)\rangle$에 제어를 걸고 $|x\rangle$에 +1을 적용한다
  - $|x\rangle$의 분포는 $|w_1(x) \oplus w_2(x)\rangle$가 모두 0이거나 모두 1인 경우에만 균일함

사이먼 합동 판별 양자 알고리즘 - 3

사이먼 합동 판별 회로 - 2



알고리즘 2: 사이먼 합동 판별

1. 2.~3.번을 $O(N)$번 반복한다.
2. 사이먼 합동 판별 회로를 구동하여 결과값을 관측한다.
3. 결과값이 둘 다 $|0\rangle$이 아닌 경우, w_1과 w_2는 합동이 아니라고 출력한다.
4. w_1과 w_2가 합동이라고 출력한다.

정리 1

문자열 w_1과 w_2, 그리고 숫자 k가 주어졌을 때,
알고리즘 2는 w_1과 w_2가 사이먼 합동을 만족할 경우 **항상 옳은 결과**를 출력하며, w_1과 w_2가 사이먼 합동이 아닐 경우 **적어도 $^1/_2$의 확률**로 옳은 결과를 출력한다.

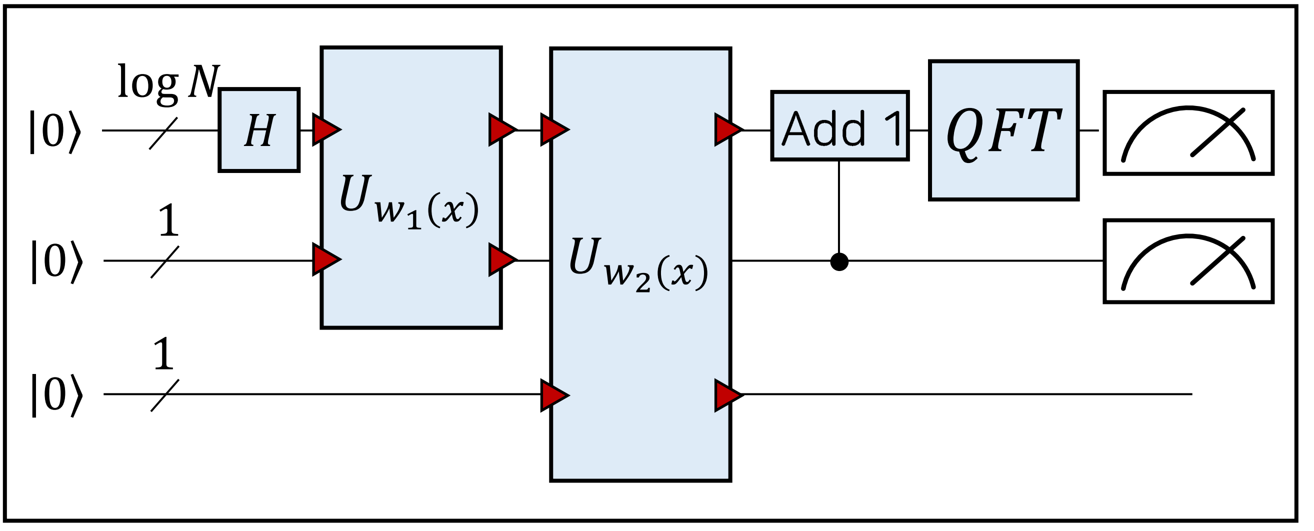
증명 스케치

- $d_k(w_1, w_2) \geq {}^N/_2$일 경우, $\Pr(|w_1(x) \oplus w_2(x)\rangle = |1\rangle) \geq {}^1/_2$
- $0 < d_k(w_1, w_2) < {}^N/_2$일 경우,

$$\Pr(QFT \neq |0\rangle) \geq \frac{\left(N - \left(N - 2 + \sqrt{2}\right)^2\right)}{N^2}$$

체르노프 바운드로 확률 보장

사이먼 합동 판별 회로 전체 구조도



향후 연구 과제 및 결론

연구 결과

- 두 문자열 사이 사이먼 합동을 판별하는 양자 프레임워크 제시
- 문자열과 언어 사이, 또는 언어 간 사이먼 합동 판별에 확장이 용이할 것으로 기대

향후 연구 과제

- 제안 알고리즘의 시간복잡도 개선
- 사이먼 합동 기반 파생 문제에 해당 프레임워크 적용

발표일자: 2024.01.30